Iterative PET Image Reconstruction Using Translation Invariant Wavelet Transform.

Jian Zhou, Lotfi Senhadji, Jean-Louis Coatrieux, Limin Luo

Iterative PET Image Reconstruction Using Translation Invariant Wavelet Transform

Jian Zhou, Lotfi Senhadji, *Senior Member, IEEE* Jean-Louis Coatrieux, *Fellow, IEEE*, Limin Luo, *Senior Member, IEEE*

***Abstract*—The present work describes a Bayesian maximum a posteriori (MAP) method using a statistical multiscale wavelet prior model. Rather than using the orthogonal discrete wavelet transform (DWT), this prior is built on the translation invariant wavelet transform (TIWT). The statistical modeling of wavelet coefficients relies on the generalized Gaussian distribution. Image reconstruction is performed in spatial domain with a fast block sequential iteration algorithm. We study theoretically the TIWT MAP method by analyzing the Hessian of the prior function to provide some insights on noise and resolution properties of image reconstruction. We adapt the key concept of local shift invariance and explore how the TIWT MAP algorithm behaves with different scales. It is also shown that larger support wavelet filters do not offer better performance in contrast recovery studies. These theoretical developments are confirmed through simulation studies. The results show that the proposed method is more attractive than other MAP methods using either the conventional Gibbs prior or the DWT-based wavelet prior.**

***Index Terms*—emission tomography, image reconstruction, maximum a posteriori, translation invariant, wavelet transform, expectation-maximization.**

I. Introduction

POSITRON emission tomography (PET) image reconstruction using *maximum a posteriori* (MAP) principles can provide improved spatial resolution and noise properties. Conventional MAP algorithms control the noise behavior through the so-called image prior as well as a smoothness constraint penalizing the roughness of image estimate and reducing the noise level. The choice of prior is however the crucial point to MAP methods. Over past years, many priors have been reported, most of which are related to the Markov random fields (MRF) that model image features (e.g., lines and edges) by means of local neighborhood structures.

While MRF priors have attracted considerable interests in image reconstruction, MRF-based approaches are typically limited to modeling very local interactions in images. In comparison to MRF priors, multiscale schemes can improve modeling and provide long range interactions, i.e., both global and local interactions at different scales can be dealt with. Previous works related to multiscale priors are primarily focused on the use of wavelets. Wavelet-based tomography reconstruction methods [1]–[13] can be divided into two categories: the analytical inversion and the iterative image reconstruction. One typical analytical method has relied on the wavelet-vaguelette decomposition (WVD) [14] that computes image reconstruction by a direct wavelet inverse from Radon domain (see [6], [7], [10] for example). Such method can offer a better image reconstruction over the conventional filtered back-projection (FBP) method, but fails to take the physical properties of PET into account. In addition, it is usually difficult to ensure the emission nonnegativity constraint when the reconstruction is performed in the wavelet domain.

To overcome these problems, statistical iterative MAP methods, coupled with a multiscale wavelet prior, have been explored. Wu *et al* [1] has early reported a wavelet prior using the space variant simultaneous autoregressive (SAR) processes. The wavelet coefficients were assumed to be correlated, and extracted using discrete wavelet transform (DWT) and anisotropic diffusion. Nowak and Kolaczyk [4] reported another multiscale approach, which however is limited to the Haar wavelet transform. Frese *et al* [9] developed the so-called wavelet graph model based on the conditional distribution of wavelet coefficients for each scale provided the information up to all coarser scales. This conditional distribution was computed through a learning stage and the classification of a specific training data set. In our previous work [13], we also proposed a wavelet-based MAP-EM (expectation-maximization) algorithm. Our multiscale prior model is mainly motivated by the fact that wavelet coefficients of most images (including medical images) are independently distributed according to generalized Gaussian distribution laws [15]–[18]. This, unlike other methods mentioned above, leads to a decorrelated multiscale wavelet prior model. By properly using the expectation-maximization (EM) algorithm, the image reconstruction was done efficiently with an iterative wavelet coefficient thresholding.

Despite the success of a wavelet prior, room for improvement remains. One challenge is due to the translation noninvariance of DWT that makes difficult a formal analysis of noise behavior as shown in [13] (we will revisit it in this paper). The present paper can be seen as an extension of our previous work with the translation invariant wavelet transform (TIWT) instead of DWT. Image TIWT representations are

This work was supported by the National Basic Research Program of China under Grant No. 2003CB718102, and a postdoctoral grant from INSERM.

J. Zhou is with the INSERM U642, Rennes F-35000, France; and the Université de Rennes 1, LTSI, Rennes F-35000, France; and the Centre de Recherche en Information Biomedicale Sino-Francais (CRIB), Rennes F-35000, France (e-mail: jian.zhou@univ-rennes1.fr).

L. Senhadji and J. L. Coatrieux are with the INSERM U642, Rennes F-35000, France, and the Université de Rennes 1, LTSI, Rennes F-35000, France; and the Centre de Recherche en Information Biomedicale Sino-Francais (CRIB), Rennes F-35000, France (e-mail: lotfi.senhadji@univ-rennes1.fr, jean-louis.coatrieux@univ-rennes1.fr).

L. Luo is with the Laboratory of Image Science and Technology (LIST), Southeast University, China; and the Centre de Recherche en Information Biomedicale Sino-Francais (CRIB), Rennes F-35000, France (e-mail: luo.list@seu.edu.cn).

often overcomplete representations, which can also be viewed as tight frames often used in domain of image processing such as denoising and deconvolution (see [19]–[23] for some specific examples). We still use the multiscale generalized Gaussian prior model. Though we have already mentioned in [13] a similar TIWT-based MAP reconstruction method, the related optimization (namely, the iterative thresholding algorithm as well [18]) with respect to TIWT was heuristic (such as the management of nonnegativity constraint) and the convergence was not addressed. Rather than using wavelet domain optimization, in this paper the reconstruction is performed in the spatial domain by means of a block EM iterative algorithm [24] which has fast and global convergent properties. The proposed method is developed over the standard discrete dyadic wavelet transform and thus is suitable for a large variety of wavelets. A theoretical study on the performance of TIWT prior model based on the local shift invariance is also proposed. This helps us to understand the main difference between the proposed multiscale wavelet prior and some conventional prior models.

This paper is organized as follows. In Section 2, we describe the TIWT-based PET reconstruction method together with the design and analysis of wavelet prior, and the optimization algorithm. Experiments are conducted in Section 3 on simulation data in order to highlight the performance of the method when compared to other approaches. The results are then discussed in Section 4 followed by some perspective issues.

II. Method

A. MAP reconstruction

In this paper, we consider the two-dimensional PET image reconstruction. Assume that the emission distribution can be discretized into p pixels with nonnegative emission rates $\boldsymbol{x} = [x_1, \ldots, x_p]' \in \mathbb{R}_+^p$ where the prime denotes matrix or vector transpose. Let y_i denote the number of emissions detected by the ith of N detectors. We consider the standard Poisson model, i.e., y_i's are independent Poisson random variables with mean

$$\bar{y}_i(\boldsymbol{x}) = [\boldsymbol{A}\boldsymbol{x}]_i + r_i \tag{1}$$

for $i = 1, \ldots, N$, where $\boldsymbol{A} := \{a_{ij}\}$ is the system matrix. r_i accounts for the presence of scatter and randoms in the data. According to this measurement model, the log-likelihood function is as follows:

$$\begin{aligned} L(\boldsymbol{y}|\boldsymbol{x}) &= \sum_i \left\{-\bar{y}_i(\boldsymbol{x}) + y_i \log \bar{y}_i(\boldsymbol{\lambda})\right\} \\ &= \sum_i \left\{-([\boldsymbol{A}\boldsymbol{x}]_i + r_i) + y_i \log([\boldsymbol{A}\boldsymbol{x}]_i + r_i)\right\} \end{aligned} \tag{2}$$

with $\boldsymbol{y} = [y_1, \ldots, y_N]' \in \mathbb{R}_+^N$ denoting the collection of measurement data.

Maximizing $L(\boldsymbol{y}|\boldsymbol{x})$ with respect to $\boldsymbol{x}$ yields the so-called maximum likelihood (ML) estimation. Recently, the Bayesian MAP estimation becomes more attractive as an alternative to the conventional ML estimation. In the MAP estimation, the prior information can be specified by a probability density on $\boldsymbol{x}$ combined with the information contained in $\boldsymbol{y}$ to produce an estimate of the unknown image. Many priors have the Gibbs form

$$\Pr(\boldsymbol{x}) \propto \exp\left\{-\beta U(\boldsymbol{x})\right\} \tag{3}$$

where $U(\boldsymbol{x})$ is the energy function, β is a smoothing parameter that controls the noise properties of the reconstructed image. Combining the likelihood function and the image prior, the MAP reconstruction is found as

$$\hat{\boldsymbol{x}}(\boldsymbol{y}) = \arg\max_{\boldsymbol{x}\geq 0} \left\{L(\boldsymbol{y}|\boldsymbol{x}) - \beta U(\boldsymbol{x})\right\}. \tag{4}$$

B. Translation invariant wavelet transform and wavelet prior

In this paper, a statistical multiscale wavelet prior model is used. This prior introduces the *a priori* knowledge on the wavelet coefficients that can be obtained through wavelet transform. We are particularly interested in the translation invariant wavelet transform. One typical TIWT is the dyadic wavelet transform where the scale is sampled along a dyadic sequence. Because we are dealing with discrete images, we further limit our attention on the discrete dyadic wavelet transform. The discrete TIWT is often carried out by the classical *algorithme à trous* (originated from [25], see also [26]–[28] or the review in [21]). This algorithm, similar to the standard DWT, can be implemented with the filter bank using pairs of low and high pass finite impulse response (FIR) wavelet filters denoted by $h_0[n]$ and $g_0[n]$ ($n \in \mathbb{Z}$) respectively. To clarify it, let us consider the case of a 2D image. It is possible to obtain the TIWT of images by a separable representation, i.e., by one-dimensional filtering of rows and columns of the image. A one-scale TIWT of image $\boldsymbol{x}$ (redefined by $\boldsymbol{x} := \boldsymbol{\theta}_0^0$) has been schematically illustrated in Fig. 1. We may view the coefficient image either as a vector in space $\mathbb{R}^p$ or as a two-dimensional discrete representation of $J \times J$ pixels. The transform has no downsampling operations for output low level images, while the used FIRs should be upsampled by a dyadic factor according to scale changes from high to low. Such scheme also requires a relatively low computational complexity. It has been shown in [21] that the implementation of TIWT with a filter bank leads to a computational complexity of $\mathcal{O}((K_h + K_g)p \log_2 p)$ where K_h and K_g are the number of non-zero samples of the initial filters h_0 and g_0 respectively.

For a M-scale TIWT, we can apply the above scheme repeatedly to the output approximation coefficients, and obtain a set of coefficient images: $\left\{\boldsymbol{\theta}_M^0, \{\boldsymbol{\theta}_m^d\}\right\}, m = 1, \ldots, M, d = 1, 2, 3$, where $\boldsymbol{\theta}_M^0 = [\theta_{M,1}^0, \ldots, \theta_{M,p}^0]' \in \mathbb{R}^p$ represents the approximation coefficients at the lowest scale M, and $\boldsymbol{\theta}_m^d = [\theta_{m,1}^d, \ldots, \theta_{m,p}^d]' \in \mathbb{R}^p$, $d = 1, 2, 3$, are the mth scale detail coefficients in three orientations (horizontal, vertical and diagonal, respectively). Note that the practical implementation of multiscale decomposition should involve the boundary conditions for the two-dimensional spatial convolution operations. For example, the periodic boundary conditions are often used. Because there are no downsampling operations, the output coefficient images always have the same size as the image being decomposed.

One critical component of our MAP reconstruction method is the use of a statistical multiscale wavelet prior model. As pointed out by several authors [29], [15], [30], [21],

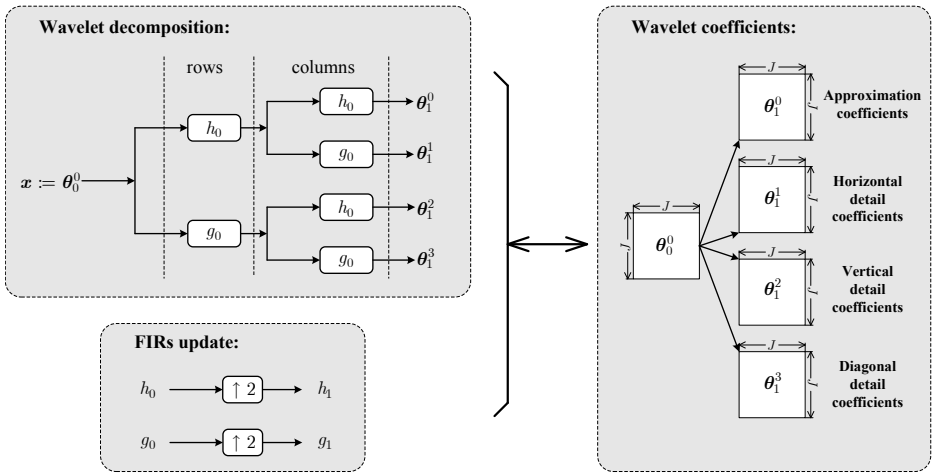


Fig. 1. One-level translation invariant wavelet transform using the algorithm *à trous*. Here $\uparrow 2$ means the upsampling operation with a factor of 2. $\boldsymbol{\theta}_m^d$'s are supposed to be square images with size $J \times J = p$. Note that the output coefficients have no downsampling operation.

[16], [17], the wavelet transform can be interpreted as a multiscale edge detector representing the singularity content of the image at multiple scales at different orientations. If the singularity is within the support of a wavelet basis function, the corresponding wavelet coefficient is large. Hence the wavelet coefficients at the singularity location tend to be large. Likewise, a smooth image region is represented by a cascade of small wavelet coefficients across scale. Since the spatial structure of many images typically consists of smooth areas separated with edges, the distribution wavelet coefficients should be sharply peaked around zero, due to the contribution of smooth areas, and have broad tails representing the contribution of the edges. It thus can be described by the generalized Gaussian distribution which has the form

$$\Pr(\boldsymbol{\theta}) \propto \exp\Big\{ -\beta\Big(\sum_k \varphi(\alpha_{M,k}^0 \theta_{M,k}^0) + \sum_m \sum_d \sum_k \varphi(\alpha_{m,k}^d \theta_{m,k}^d)\Big)\Big\} \tag{5}$$

where $\varphi(t) = |t|^s$ with $0 < s \leq 2$, and $\alpha_{m,k}^d$ represents the weight for the kth coefficient at the dth orientation and the mth scale. Note that the tail of this distribution becomes increasingly heavy as s approaches zero. For real cases, the optimal s is seldom known to us, this paper only considers the simple probabilistic model by taking $s = 1$, which relates to the familiar Laplacian distribution. The Laplacian distribution has proven to be useful for the modeling of heavy-tailed wavelet coefficients [15]. Unlike the Gaussian distribution, it penalizes less image features. Therefore, one can expect that a MAP with such prior can produce edge-preserved image reconstruction.

The specification of α's for each coefficient is of importance for a multiscale wavelet prior model. One can emphasize the role of some specific wavelet coefficients by increasing the corresponding α's. Nevertheless, this is not realistic due to the lack of *a priori* information. A useful choice can be made by considering the approximate energy relationship of coefficient images across various scales. In Appendix, we have derived the two-scale Fourier domain coefficient relationship which indicate that the total energy of a higher level approximation coefficient image can be divided into four parts corresponding to four lower level coefficient images. The energy proportion relates to the wavelet filters by a constant factor $1/C^2$ ($C \neq 1$). Therefore, the energy of wavelet coefficients is scale dependent. To remove this dependence, we suggest the following α's

$$\alpha_{m,k}^d := \alpha_m = C^{-2m}, \quad C > 0. \tag{6}$$

By using these weights, we then obtain equally contributing coefficients from the viewpoint of balanced signal energy. Of course, (6) is only one possible choice for α's, a thorough discussion on weight modeling can be found in [16]. Recently, the choice of parameter α's has been shown of relevance for the Besov space norm [30], [31].

It is worth to note that $\boldsymbol{\theta}_M^0$ serves as a low-level approximation of the original image $\boldsymbol{x}$, while each $\boldsymbol{\theta}_{m,d}$ (for $m = 1, \ldots, M-1$ and $d = 1, 2, 3$) depicts the local image feature. Thus, both the global and local image *a priori* information has been included by the proposed multiscale wavelet prior. This differs from the conventional MRF-based prior where information about local features is included only. It is easy to show that such global prior gives preference to image reconstructions with smaller norms (when $s = 1$, the function φ is the l_1 norm). Thus it plays a similar role of noise suppression, and then can be expected to yield additional improvement in image reconstruction. Finally, the energy function used for our MAP estimation (4) is

$$U_{\mathrm{WV}}(\boldsymbol{x}) = \sum_k \varphi(\alpha_M \theta_{M,k}^0) + \sum_m \sum_d \sum_k \varphi(\alpha_m \theta_{m,k}^d). \tag{7}$$

C. Prior model study using local shift invariance

In PET imaging, it is often important to quantify a MAP reconstruction method in terms of resolution and variance of the resulting images. These measures can be used in comparing different reconstruction algorithms on a particular imaging system. Due to the nonlinear MAP estimator, a direct computation of these quantities is nontrivial. Early works by [32], [33] including [34], [35] have used the approximate

expressions based on the implicit function theorem, the Taylor expansion and the chain rule. Typically, the jth local impulse response has the form

$$l_j(\hat{\boldsymbol{x}}) \approx (\boldsymbol{F} + \beta \boldsymbol{U}(\hat{\boldsymbol{x}}))^{-1} \boldsymbol{F} \boldsymbol{e}_j \tag{8}$$

and the variance is

$$\mathsf{Var}_j(\hat{\boldsymbol{x}}) \approx (\boldsymbol{e}_j)'(\boldsymbol{F} + \beta \boldsymbol{U}(\hat{\boldsymbol{x}}))^{-1} \boldsymbol{F} (\boldsymbol{F} + \beta \boldsymbol{U}(\hat{\boldsymbol{x}}))^{-1} \boldsymbol{e}_j. \tag{9}$$

where $\boldsymbol{e}_j$ is the jth unit vector, $\boldsymbol{F} := \boldsymbol{A}' D[q_i] \boldsymbol{A}$ is the Fisher information matrix given the measurement data; $D[q_i]$ denotes a diagonal matrix with q_i (representing the reciprocal of the variance of y_i) being the ith diagonal entry. q_i's are usually solved by the *data-plugin* techniques (refer to [35]–[37] for more details). $\boldsymbol{U}(\hat{\boldsymbol{x}})$ is the Hessian matrix of the energy function U. Provided the measured data and system, the Fisher information may not change anymore. Hence, it is the Hessian $\boldsymbol{U}(\hat{\boldsymbol{x}})$, varying with the energy function as well as the used prior model, that makes one MAP method distinguishing from another. While the incorporation of image prior may reduce the noise variance, it would lead to a bias in image estimation. By assuming the local shift invariance for both Fisher and Hessian matrices, it further can be shown [38], [39] that

$$l_j(\hat{\boldsymbol{x}}) \approx \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{\boldsymbol{F}\boldsymbol{e}_j\}}{\mathcal{F}\{\boldsymbol{F}\boldsymbol{e}_j\} + \beta \mathcal{F}\{\boldsymbol{U}(\hat{\boldsymbol{x}})\boldsymbol{e}_j\}} \right\} \tag{10}$$

and

$$\mathsf{Var}_j(\hat{\boldsymbol{x}}) \approx \sum_n \left\{ \frac{[\mathcal{F}\{\boldsymbol{F}\boldsymbol{e}_j\}]_n}{([\mathcal{F}\{\boldsymbol{F}\boldsymbol{e}_j\}]_n + \beta[\mathcal{F}\{\boldsymbol{U}(\hat{\boldsymbol{x}})\boldsymbol{e}_j\}]_n)^2} \right\} \tag{11}$$

where $\mathcal{F}\{\cdot\}$ represents the 2D discrete Fourier transform operator, $\mathcal{F}^{-1}\{\cdot\}$ is the corresponding inverse, and the division in (10) is an element-by-element division. These two approximations enable us to identify the principal differences of MAP methods by analyzing the locally spatial response $\boldsymbol{U}(\hat{\boldsymbol{x}})\boldsymbol{e}_j$.

To make it clear, let us first consider the conventional MRF priors. The Gibbs energy of most conventional MRF priors can be written in the form [40]

$$U_{\mathrm{MRF}}(\boldsymbol{x}) = \sum_j w_j \phi([\boldsymbol{C}\boldsymbol{x}]_j) \tag{12}$$

where ϕ is the potential function, $\boldsymbol{C}$ denotes the difference operator, and w_j's are weighting parameters. For a usual second-order MRF, we may partition $\boldsymbol{C}$ into four parts: $\boldsymbol{C} = [\boldsymbol{C}_1', \boldsymbol{C}_2', \boldsymbol{C}_3', \boldsymbol{C}_4']'$ where $\boldsymbol{C}_\ell$'s ($\ell = 1, \ldots, 4$) are shift invariant operations representing the horizontal, vertical and two diagonal difference operations respectively. The finite difference operation can be performed by using two-dimensional convolution. Typically, the two-dimensional convolution kernels of $\boldsymbol{C}_\ell$'s can be [41]:

$$\begin{bmatrix} -1 & 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } \begin{bmatrix} 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix},$$

respectively. The Hessian of $U_{\mathrm{MRF}}(\boldsymbol{x})$ is therefore

$$\boldsymbol{U}_{\mathrm{MRF}}(\boldsymbol{x}) = \sum_\ell (\boldsymbol{C}_\ell)' D[\xi_j^\ell(\boldsymbol{x})] \boldsymbol{C}_\ell \tag{13}$$

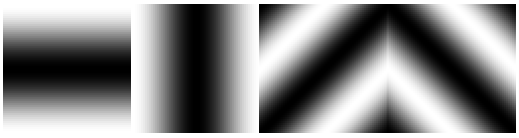


Fig. 2. Four frequency magnitude response images of the kernel of a usual second-order MRF prior. From left to right: $\ker(\boldsymbol{C}_1'\boldsymbol{C}_1)$ (horizontal), $\ker(\boldsymbol{C}_2'\boldsymbol{C}_2)$ (vertical), $\ker(\boldsymbol{C}_3'\boldsymbol{C}_3)$, and $\ker(\boldsymbol{C}_4'\boldsymbol{C}_4)$ (two diagonals), respectively. Note that the DC component of Fourier transform is located at the center of each image.

where $\xi_j^\ell(\boldsymbol{x}) = w_j \ddot{\phi}([\boldsymbol{C}_\ell \boldsymbol{x}]_j)$ with $\ddot{\phi}$ the second order derivative of ϕ. If we assume that the $\xi_j^\ell(\boldsymbol{x})$'s are spatially stationary, we can obtain the following approximation:

$$\begin{aligned} \boldsymbol{U}_{\mathrm{MRF}}(\boldsymbol{x})\boldsymbol{e}_j &\approx \sum_{\ell=1}^{4} \underline{\xi}_j^\ell(\boldsymbol{x}) (\boldsymbol{C}_\ell)' \boldsymbol{C}_\ell \boldsymbol{e}_j \\ &\Leftrightarrow \sum_{\ell=1}^{4} \underline{\xi}_j^\ell(\boldsymbol{x}) \ker(\boldsymbol{C}_\ell'\boldsymbol{C}_\ell). \end{aligned} \tag{14}$$

The above equation shows the linear combination of high-pass filters $\{\ker(\boldsymbol{C}_\ell'\boldsymbol{C}_\ell)\}_{\ell=1,\ldots,4}$ (where $\ker(\boldsymbol{X})$ means the two-dimensional convolution kernel of a shift invariant operator $\boldsymbol{X}$) with portions adjusted by $\underline{\xi}_j^\ell(\boldsymbol{x})$'s (it is possible to use the approximate exchange to compute $\underline{\xi}_j^\ell(\boldsymbol{x})$'s, see Eq. 16 in [40] for instance). This indicates that the influence of a MRF prior affects on the higher frequency domain. Fig. 2 shows the magnitude response of the corresponding filters, where different frequency concentrations (i.e., $\{[-\pi, -\pi/2] \cup [\pi/2, \pi]\} \times \{[-\pi, -\pi/2] \cup [\pi/2, \pi]\}$) and orientations can be well recognized. The coefficients $\underline{\xi}_j^\ell(\boldsymbol{x})$'s control the local behavior of each filter. For most of the nonlinear potential functions, we see that $\underline{\xi}_j^\ell(\boldsymbol{x})$'s are signal dependent, which leads to the adaptive image reconstruction (e.g., the edge-preserved filtering). When the coefficients do not change largely (usually for pixels within relatively smooth regions), the effect of such prior is proportional to a conventional quadratic prior. In fact, $\underline{\xi}_j^\ell(\boldsymbol{x})$'s in this case is also independent of position j. Thus, we may have

$$\boldsymbol{U}_{\mathrm{MRF}}(\boldsymbol{x})\boldsymbol{e}_j \propto \sum_{\ell=1}^{4} \ker(\boldsymbol{C}_\ell'\boldsymbol{C}_\ell) = \ker \begin{pmatrix} -2 & -1 & -2 \\ -1 & 12 & -1 \\ -2 & -1 & -2 \end{pmatrix}$$

as a constant high-pass filter kernel.

As for the proposed wavelet prior, we employ the same strategy by separating $\boldsymbol{W}$ into a set of operators: $\{\boldsymbol{W}_M^0, \{\boldsymbol{W}_m^d\}\}$, $m = 1, \ldots, M$, $d = 1, \ldots, 3$ where $\boldsymbol{W}_m^d$ represents the partial wavelet transform that acts on $\boldsymbol{x}$ to produce the coefficients $\boldsymbol{\theta}_m^d$, i.e., $\boldsymbol{\theta}_m^d = \boldsymbol{W}_m^d \boldsymbol{x}$. Then, according to (7), the Hessian of the proposed multiscale wavelet prior can be expressed as

$$\begin{aligned} \boldsymbol{U}_{\mathrm{WV}}(\boldsymbol{x}) = &(\boldsymbol{W}_M^0)' D[\lambda_{M,k}^0(\boldsymbol{x})](\boldsymbol{W}_M^0) \\ &+ \sum_{m=1}^{M} \sum_{d=1}^{3} (\boldsymbol{W}_m^d)' D[\lambda_{m,k}^d(\boldsymbol{x})](\boldsymbol{W}_m^d) \end{aligned} \tag{15}$$

with $\lambda_{m,k}^d(\boldsymbol{x}) = \alpha_m^2 \ddot{\varphi}(\alpha_m [\boldsymbol{W}_m^d \boldsymbol{x}]_k)$ and $\ddot{\varphi}$ the second order derivative of φ. Using the similar assumption that the

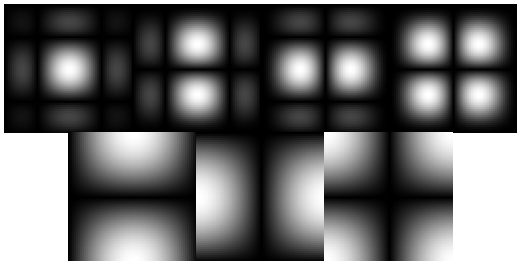


Fig. 3. Frequency magnitude response images involved by a two-level TIWT prior with Haar wavelet filters. The first row (from left to right), the second level: $\ker((\boldsymbol{W}_2^0)'(\boldsymbol{W}_2^0))$ (low-pass filter), $\ker((\boldsymbol{W}_2^1)'(\boldsymbol{W}_2^1))$ (horizontal), $\ker((\boldsymbol{W}_2^2)'(\boldsymbol{W}_2^2))$ (vertical), $\ker((\boldsymbol{W}_2^3)'(\boldsymbol{W}_2^3))$ (diagonal); The second row (from left to right), the first level: $\ker((\boldsymbol{W}_1^1)'(\boldsymbol{W}_1^1))$ (horizontal), $\ker((\boldsymbol{W}_1^2)'(\boldsymbol{W}_1^2))$ (vertical), $\ker((\boldsymbol{W}_1^3)'(\boldsymbol{W}_1^3))$ (diagonal). Note that the DC component of Fourier transform is located at the center of each image.

$\lambda_{m,k}^d(\boldsymbol{x})$'s vary slowly, we have

$$\begin{aligned}\boldsymbol{U}_{\mathrm{WV}}(\boldsymbol{x})\boldsymbol{e}_j &\approx \underline{\lambda}_{M,j}^0(\boldsymbol{x})(\boldsymbol{W}_M^0)'(\boldsymbol{W}_M^0)\boldsymbol{e}_j \\ &\quad + \sum_{m=1}^{M}\sum_{d=1}^{3}\underline{\lambda}_{m,j}^d(\boldsymbol{x})(\boldsymbol{W}_m^d)'(\boldsymbol{W}_m^d)\boldsymbol{e}_j \\ &\Leftrightarrow \underline{\lambda}_{M,j}^0(\boldsymbol{x})\ker((\boldsymbol{W}_M^0)'(\boldsymbol{W}_M^0)) \\ &\quad + \sum_{m=1}^{M}\sum_{d=1}^{3}\underline{\lambda}_{m,j}^d(\boldsymbol{x})\ker((\boldsymbol{W}_m^d)'(\boldsymbol{W}_m^d)).\end{aligned} \tag{16}$$

$\boldsymbol{U}_{\mathrm{WV}}(\boldsymbol{x})\boldsymbol{e}_j$ is no more than $\boldsymbol{U}_{\mathrm{MRF}}(\boldsymbol{x})\boldsymbol{e}_j$, having the same form of linear filter combination with signal related coefficients $\underline{\lambda}_{m,j}^d(\boldsymbol{x})$'s. However, these filters consist of a low-pass filter $\ker((\boldsymbol{W}_M^0)'(\boldsymbol{W}_M^0))$, and several subband high-pass filters $\{\ker((\boldsymbol{W}_m^d)'(\boldsymbol{W}_m^d))\}$, $m = 1,\ldots,M$, $d = 1,\ldots,3$. Note that for each scale, there is only one diagonal filter while for the MRF prior two separable diagonal filters are used. The frequency concentration of high-pass filters are scale dependent. As an example, Fig. 3 shows the frequency response of the Haar wavelet filters up to scale $M = 2$. Typically, it is concentrated on $\{[-2^{-m+1}\pi, -2^{-m}\pi] \cup [2^{-m}\pi, 2^{-m+1}\pi]\} \times \{[-2^{-m+1}\pi, -2^{-m}\pi] \cup [2^{-m}\pi, 2^{-m+1}\pi]\}$, $m = 1,\ldots,M$. Here, let us consider the one-scale Haar wavelet prior in which we slightly ignore the global regularization as well as the low-level approximation wavelet coefficients. We suppose the signal related coefficients $\underline{\lambda}_{m,j}^d(\boldsymbol{x})$'s do not vary too much so that they are independent of both signal content and pixel locations, it can be shown that

$$\begin{aligned}\boldsymbol{U}_{\mathrm{WV}}(\boldsymbol{x})\boldsymbol{e}_j &\propto \sum_{d=1}^{3}\ker((\boldsymbol{W}_1^d)'(\boldsymbol{W}_1^d)) \\ &= \ker\begin{pmatrix} -2 & -1 & -2 \\ -1 & 12 & -1 \\ -2 & -1 & -2 \end{pmatrix}\end{aligned} \tag{17}$$

This implies that the conventional MRF prior can be viewed as a special case of Haar wavelet prior. When $M > 1$, this clearly yields larger range of frequency domain (in comparison with MRF prior) and thus more signal features could be involved. From this point of view, the wavelet prior is usually different from the MRF prior.

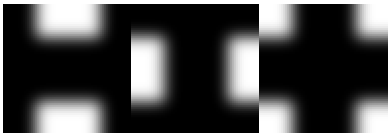


Fig. 4. Frequency magnitude response images of high-pass filters involved by a one-level TIWT prior with 'db4' wavelet filters. From left to right: $\ker((\boldsymbol{W}_1^1)'(\boldsymbol{W}_1^1))$ (horizontal), $\ker((\boldsymbol{W}_1^2)'(\boldsymbol{W}_1^2))$ (vertical), $\ker((\boldsymbol{W}_1^3)'(\boldsymbol{W}_1^3))$ (diagonal). Note that the DC component of Fourier transform is located at the center of each image.

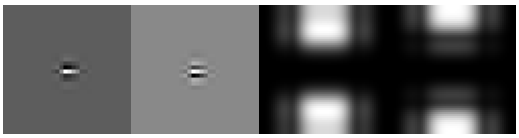


Fig. 5. The first two images show two kernels extracted from the operator $(\boldsymbol{W}_1^1)'(\boldsymbol{W}_1^1)$ where $\boldsymbol{W}_1^1$ computed by the standard DWT with db4 wavelet filters. For good visualization, only center part of images (32×32) is given. The last two images are the corresponding frequency magnitude responses. Note that the DC component of Fourier transform is located at the center of each image.

Similarly, let us check different wavelet filters. In Fig. 4, we give out another set of filters computed with Daubechies wavelet filters of length 8 (denoted by 'db4'), i.e., with larger support wavelet basis. Clearly, the frequency orientation of such wavelet filters are the same as the Haar wavelet. Nevertheless, the high-pass effect of the related filters becomes stronger (i.e., the 'db4' filters rolls off more quickly around the cutoff frequency than the Haar filters. Note that the brighter color in images indicate higher magnitudes), and therefore high frequency components will be more penalized. Since the Hessian occurs only in the denominator of (10) and (11), the consequence is that: the image variance could be reduced, leading to smoother reconstructions as compared to a Haar wavelet prior; On the other hand, the image bias increases due to the large penalization of the local impulse response. Thus, the performance of TIWT MAP changes with different wavelet filters. In our experiments, we will show that this method for the particular application of lesion contrast recovery is actually degraded if large support wavelets are used.

It is worth to note that although the above local analysis may not hold for DWT-based MAP method since DWT is shift-varying, it is still possible to get some insights on the behavior of a shift-varying wavelet prior. In Fig. 5, we studied a prior computed by the standard DWT. These two kernels correspond to two consecutive rows extracted from the operator $(\boldsymbol{W}_1^1)'(\boldsymbol{W}_1^1)$. Both of them have been shifted properly to the center of image. Different spatial profiles and frequency magnitudes illustrate the shift-varying property. Notice also that the spatial profiles are asymmetric, which indicates that the local impulse response is anisotropic and even irregular. This may complicate the analysis of noise behavior in image reconstruction. We will address this issue in Section 3.

D. Optimization

We have adapted the block sequential regularized expectation maximization (BSREM) algorithm [24] to compute the MAP reconstruction. This algorithm is an extension of

RAMLA (row-action maximum likelihood algorithm) [42] for MAP reconstruction. It has proven to be fast and to have guaranteed convergence properties. In this algorithm, a full iteration consists of a set of subiterations, each being performed sequentially using one predetermined block of measurement data. Let us define a partition of the integer set $\Omega = [1, 2, \ldots, N]$ into B disjoint subsets S_b, such that $\bigcup_{b=1}^{B} S_b = \Omega$. Then, the BSREM iteration scheme is given as follows:

1. Setup $\hat{\boldsymbol{x}}^0$, for $n = 0, 1, \ldots$, $\hat{\boldsymbol{x}}^n = \hat{\boldsymbol{x}}^{n,0}$, and for $b = 1, \ldots, B$ and $j = 1, \ldots, p$
$$\hat{x}_j^{n,b} = \hat{x}_j^{n,b-1} + \rho^n \hat{x}_j^{n,b-1} \sum_{i \in S_b} a_{ij} \left(\frac{y_i}{\sum_{d=1}^{p} a_{id} \hat{x}_d^{n,b-1} + r_i} - 1 \right);$$
2. Set $\hat{\boldsymbol{x}}^{n+1/2} \equiv \hat{\boldsymbol{x}}^{n,B}$ and define for $j = 1, \ldots, p$
$$\hat{x}_j^{n+1} = \hat{x}_j^{n+1/2} - \rho^n \hat{x}_j^{n+1/2} \left[\nabla U_{\mathrm{WV}}(\hat{\boldsymbol{x}}^{n+1/2}) \right]_j;$$
3. Apply the nonnegativity constraint:
$$\hat{x}_j^{n+1} = \max\left\{ \hat{x}_j^{n+1}, \tau \right\} \quad \forall j = 1, \ldots, p.$$

Here, $\{\rho^n\}$ is a sequence of positive relaxation parameters such that $\lim_{n\to\infty} \rho^n = 0$ and $\sum_{n=0}^{\infty} \rho^n = \infty$. $\tau > 0$ is a threshold, compensating the nonnegativity constraint. $\nabla U_{\mathrm{WV}}(\boldsymbol{x}) = [\frac{\partial U_{\mathrm{WV}}}{\partial x_1}, \ldots, \frac{\partial U_{\mathrm{WV}}}{\partial x_p}]'$ is the gradient of function U_{WV} with respect to $\boldsymbol{x}$. By letting $\boldsymbol{\eta}_m^d = [\eta_{m,1}^d, \ldots, \eta_{m,p}^d]'$ where $\eta_{m,k}^d = \alpha_m \dot{\varphi}(\alpha_m [\boldsymbol{W}_m^d \boldsymbol{x}]_k)$ and $\dot{\varphi}$ is the first derivative of φ, then

$$\nabla U_{\mathrm{WV}}(\boldsymbol{x}) = (\boldsymbol{W}_M^0)' \boldsymbol{\eta}_M^0 + \sum_m \sum_d (\boldsymbol{W}_m^d)' \boldsymbol{\eta}_m^d. \tag{18}$$

The calculation of $\nabla U_{\mathrm{WV}}(\boldsymbol{x})$ is equivalent to the backward TIWT that maps wavelet coefficients into an image. Here we treat $\boldsymbol{\eta}_m^d$'s as wavelet coefficient images so $\boldsymbol{W}_m^d$'s play the role of backward wavelet transform, mapping $\boldsymbol{\eta}_m^d$'s back to the image space. Note that TIWT is nonorthogonal so this backward is not really the wavelet inverse transform. It can be carried out by a (backward) filter bank with the same computational complexity as the forward TIWT [21]. Note that because the function φ is nondifferentiable when $s = 1$, a smooth approximation as shown below on is often desired as computing the gradient of the energy function $U_{\mathrm{WV}}(\boldsymbol{x})$. In this paper we assume that such approximation will not affect too much the behavior of prior model, hence we expect that the BSREM would converge to an image reconstruction close to the real MAP solution. Of course, there are other optimization algorithms available, such as the interior-point method [43] and the splitting algorithm [23] (the latter does not need a differentiable energy function).

III. Simulation studies

This section provides simulation results in order to study quantitatively the performance of the TIWT-based MAP reconstruction. We considered the three-dimensional NU2–2000 National Electrical Manufacturers Association (NEMA) phantom [44]. This phantom is described on a space of 350 mm$\times 350$ mm$\times 180$ mm (axial). The body wall thickness is 3 mm, the lung is modeled as 250mm radius cylinder in the center of the phantom. It consists of 6 spherical lesions of 10 mm, 13 mm, 17 mm, 22 mm, 28 mm, and 37 mm diameters. The first four lesions are hot spheres and the last two cold spheres. In the emission phantom, the activity was applied by setting the background (i.e., the lung and other body tissues) to 2. All hot lesions had a contrast equal to 5 while cold lesions were set at 0.5. We also assumed that the background has uniform attenuation coefficient of 0.0095 mm^{-1}. The whole volume was digitized with 64×64 (axial) voxels. For the two-dimension image reconstruction, we selected the central slice (the related emission activity image is shown in Fig. 9(a)). We simulated a PET scanner operating in two-dimensional mode which produces the projection data with 80 radial bins and 64 angular views (evenly spread over $180°$). The detector efficiencies were generated by a pseudo-random log-normal variance of 0.3. The sinogram was globally scaled to a mean sum of $200,000$ true events. Pseudo-random independent Poisson variants were drawn according to (1), and a uniform field of Poisson distributed background scatter events with known mean of 10% random coincidences was also considered in our experiments.

Reconstructions of the simulated data were performed using the MRF-based and the wavelet-based MAP algorithms. Besides the proposed TIWT-based MAP, we also used the DWT for comparison purpose. For the DWT-based algorithm, the optimization was not changed. The weights α's had to be set to 1 according to the signal energy balance previously mentioned (note that in DWT there is a downsampling operation so that the extra factor as well as C resulting from filtering taps would be canceled, see [21] for more details about DWT). DWT has lower computational complexity and lower memory storage than TIWT, which can be more attractive for higher dimensional image reconstruction. For the MRF model, we chose two commonly used potential functions: $\phi(t) = |t|^2$ and $\phi(t) = |t|$. The former leads to the Gaussian quadratic MRF model while the latter is close to the well-known total variational (TV) regularization. In the following, we use the terms QUAD-MAP, TV-MAP, DWT-MAP and TIWT-MAP to identify different methods. All MAP methods were implemented by using the BSREM optimization. In BSREM, we fixed $B = 16$ blocks. Each block had the same size. The bth block index subset S_b ($\forall b = 1, \ldots, B$) was chosen according to the angular view. More precisely, let us denote n_a the number of projection angles. For the bth block, we consider those projection data in views: $\{b + 16a \mid a = 0, 1..., \text{and } (b+16a) \le n_a\}$. The threshold τ was set to 10^{-8}. Since the function φ (including the TV $\phi(t) = |t|$) is not differentiable at zero, we have used the *smooth* approximation proposed in [45], [16]: $\varphi(t) \approx \sqrt{|t|^2 + \epsilon} - \sqrt{\epsilon}$, where ϵ, the stabilization constant, is set to 10^{-6}. The relaxation parameter was generated according to [24]: $\rho^n = \rho^0/(n+1)^{0.1}$, where the initial value ρ^0 was chosen carefully for each algorithm to avoid any divergence. The initial estimate $\hat{\boldsymbol{x}}^0$ was a uniform disk with a constant emission activity. We ran 200 iterations

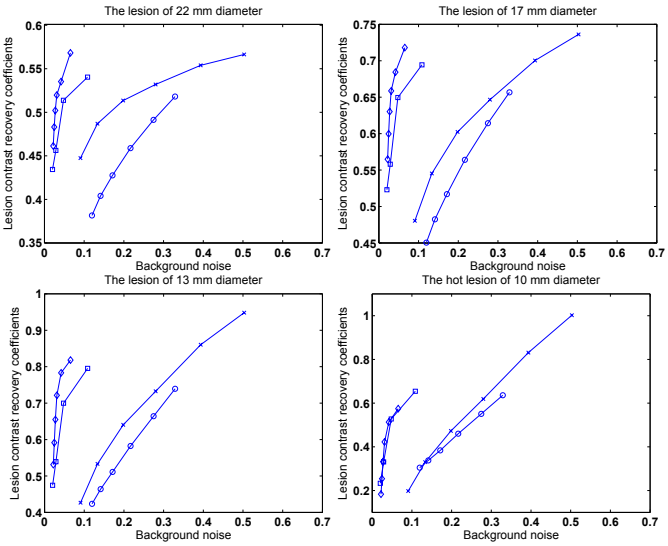


Fig. 6. Background noise standard deviation versus hot lesion contrast recovery coefficient. The reconstruction methods are: QUAD-MAP (circle), TV-MAP (square), DWT-MAP (cross), and TIWT-MAP (diamond). Both DWT and TIWT have used the Haar wavelet including a total $M = 3$ decomposition level.

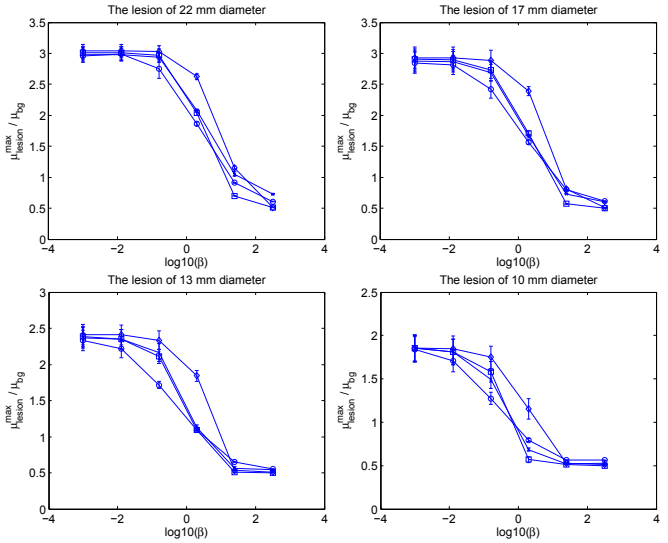


Fig. 7. The changes of uptake estimation $\mu^{\max}_{\text{lesion}}/\mu_{\text{bg}}$ as a function of β for hot lesions with different diameters. The reconstruction methods are: QUAD-MAP (circle), TV-MAP (square), DWT-MAP (cross) and TIWT-MAP (diamond).

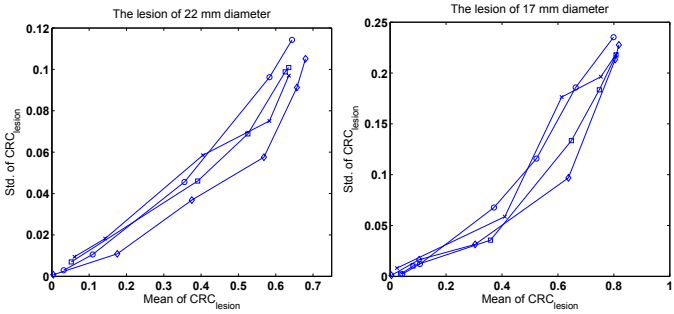
to ensure the effective convergence.

A. Lesion contrast recovery analysis

We first studied the performance of different MAP algorithms for lesion recovery by using the contrast recovery coefficients (CRC) for different sized hot lesions. The lesion CRC is defined as (see Eq. (16) in [46]):

$$\text{CRC}_{\text{lesion}} = \frac{\mathsf{E}\{\mu^{\max}_{\text{lesion}}\}/\mathsf{E}\{\mu_{\text{bg}}\} - 1}{\mu^{\text{true}}_{\text{lesion}}/\mu^{\text{true}}_{\text{bg}} - 1} \tag{19}$$

where $\mu^{\max}_{\text{lesion}}$ and μ_{bg} denotes the maximum activity of the reconstructed lesion and the mean activity background



(a) (b)

Fig. 8. The mean and standard deviation of $\text{CRC}_{\text{lesion}}$ relations of two lesion of (a) 22 mm and (b) 17 mm diameters. The reconstruction methods are: QUAD-MAP (circle), TV-MAP (square), DWT-MAP (cross) and TIWT-MAP (diamond).

respectively, and $\mu^{\text{true}}_{\text{lesion}}$ and $\mu^{\text{true}}_{\text{lesion}}$ are the corresponding true activity values. The expectation operation $\mathsf{E}\{\cdot\}$ was replaced by the ensemble mean computed from reconstructions of 200 replicate noisy sinograms. The background mean and standard deviation were calculated from 4 selected circular regions in the lung area. For the wavelet-based MAP algorithms, we have used the Haar wavelet filters for instance, the maximum decomposition level was set to $M = 3$ for both DWT and TIWT. Fig. 6 plots the contrast recovery coefficients (CRC) of four hot lesions versus the spatial standard deviation of the reconstructed background. Note that the recovery rate is related to the lesion size. For lesions of 22, 17 and 13 mm diameters, the CRCs of TIWT-MAP are higher. This shows the advantage of the multiscale TIWT prior in achieving higher contrast recovery at matched noise levels. The DWT-MAP shows better performance than QUAD-MAP while it is inferior to either TIWT-MAP or TV-MAP. For the smallest lesion, the performance of TV-MAP is close to TIWT-MAP. We analyzed the uptake $\mu^{\max}_{\text{lesion}}/\mu_{\text{bg}}$ as a function of β. The results are shown in Fig. 7 with the mean and standard deviation. It can be seen that the proposed TIWT-MAP has relatively higher uptake than other MAP methods for all hot lesions with different sizes, which again indicates better performance. Also in this particular experiment, there seems no large different performance between TV-MAP and DWT-MAP while they are better than QUAD-MAP. To perform further evaluation, we used a modified CRC by excluding the expectation operation. Then, for each image reconstruction, we had CRC's for different lesions. With 200 reconstructions in total, we were able to calculate the CRC mean and the related standard deviation. Fig. 8 shows the changes of mean CRC as a function of standard deviation for two lesions of 22 mm and 17 mm diameters respectively. These results confirm our proposed TIWT-MAP algorithm since the standard deviation of TIWT-MAP is usually lower than others for some fixed mean values. In Fig. 9, we show different MAP image reconstructions from one single noise sinogram. Here the β value was selected according to Fig. 6(d) so that all methods produced the nearly equal CRC (≈ 0.4) for the 10 mm lesion. The results yielded by TIWT-MAP, TV-MAP and DWT-MAP are visually similar with well preserved sharp edges. In this experiment, they are

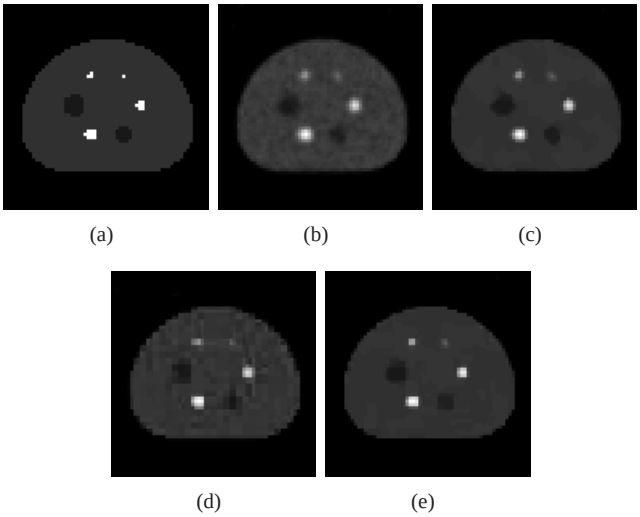


Fig. 9. (a) The simulated phantom; (b)–(e) are image reconstructions from QUAD-MAP, TV-MAP, DWT-MAP, and TIWT-MAP, respectively. The smoothing parameters for four methods were chosen to achieve approximately matched $\mathrm{CRC}_{\mathrm{lesion}}$ (≈ 0.4) for the smallest hot lesion of 10 mm diameter according to curves in Fig. 6(d).

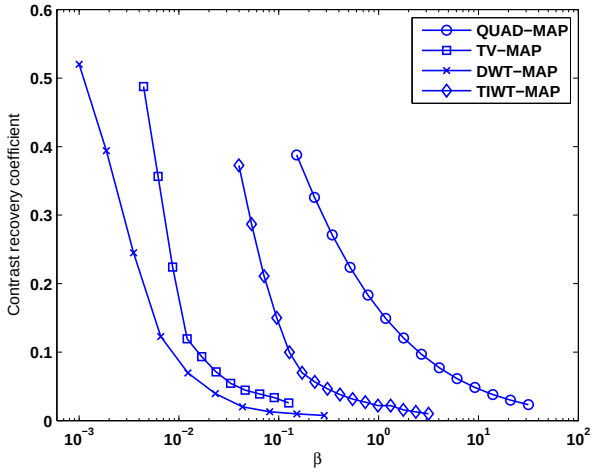


Fig. 10. Parameter β versus contrast recovery coefficient (CRC) on pixel of interest at the center of phantom for different MAP methods.

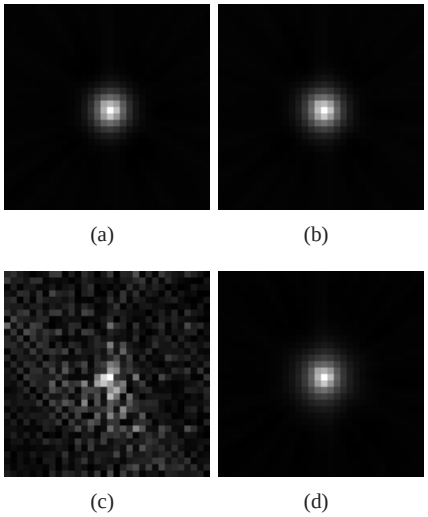


Fig. 11. Local impulse response images of four MAP methods (from (a) to (d)) QUAD-MAP, TV-MAP, DWT-MAP and TIWT-MAP, respectively. For a good visualization, only part of images (32×32) are displayed. The parameters β for different MAP methods were chosen to create a nearly same target CRC (≈ 0.05) at pixel of interest using curves shown in Fig. 10. Both DWT and TIWT were computed with Haar wavelet filters and the maximum decomposition level $M = 3$.

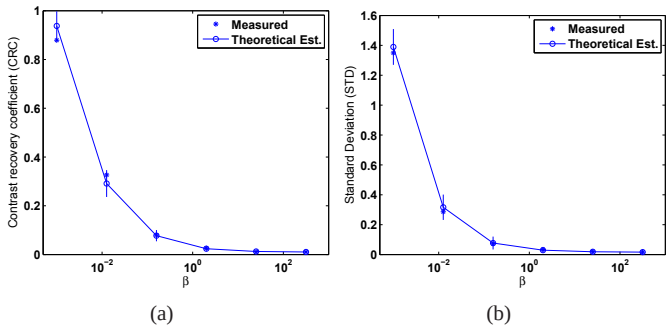


Fig. 12. (a) Comparison of the experimental CRCs and theoretical CRCs using the approximate expression versus β; (b) Comparison of the standard deviation of point of interest versus β. The measured standard deviation were computed at the center of phantom across 200 independent reconstructions.

better than that obtained by QUAD-MAP.

This simulation study can be confirmed through a theoretical analysis based on (10) and (11). The local CRC of any pixel j is often used to characterize the resolution property. The definition of local CRC is [34], [35]

$$\mathrm{crc}_j(\hat{\boldsymbol{x}}) = (\boldsymbol{e}_j)' l_j(\hat{\boldsymbol{x}}). \tag{20}$$

The β versus CRC curves are plotted in Fig. 10. This figure indicates a monotonic relationship between β and CRC for a given pixel of interest. We also studied the related local impulse response functions of different MAP methods. According to Fig. 10, we chose a target CRC (≈ 0.05) for each MAP method, and then displayed the impulse images (Fig. 11). It can be seen that these methods except DWT lead to a well-shaped point spread function. The result is somewhat consistent with our previous analysis, pointing out that the translation noninvariance of DWT causes an irregular local impulse response function.

We also evaluated the accuracy of the theoretical approximate expressions for the TIWT-MAP estimation. The theoretical CRC was compared to the experimental CRC which was measured from reconstructions of two noiseless data sets: 1) the original phantom sinogram, and 2) the sinogram of the phantom after perturbation of a single pixel of interest. Fig. 12(a) shows a comparison of the measured and theoretical predicted CRCs for that pixel of interest. The approximations are in good agreement with the measured CRCs. We then computed the pixel-wise variance from 200 reconstructions from independent data sets. The standard deviation versus parameter β curves from both measured and theoretical estimated results are shown in Fig. 12(b). Again, the agreement is good: the theoretical predictions do not show a significant increase in error.

To combine the CRC with the background noise level, we used the contrast-to-noise ratio (CNR), which has the

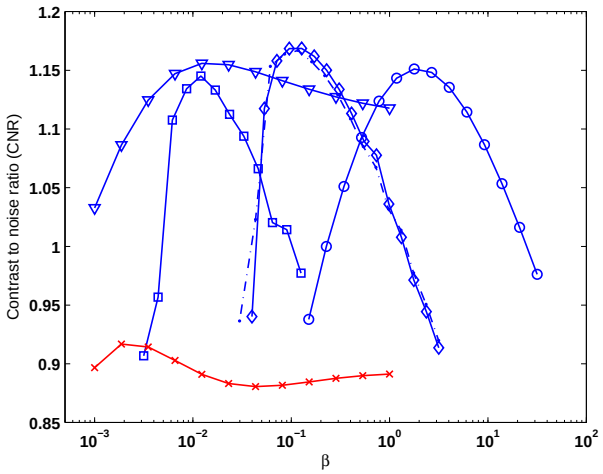


Fig. 13. The relation curves between CNR and parameter β of different MAP methods for pixel of interest located at the center of phantom: QUAD-MAP (circle), TV-MAP (square), TIWT-MAP (diamond) and the modified TIWT-MAP (dash-dot) with $\alpha_M = 0$. For DWT-MAP, we computed two curves: one is for the pixel at (33, 33) (triangle) and the other is for (32, 33) (cross). Note that the performance of DWT-MAP changes greatly as we only shifted one pixel spatial location.

form [34]

$$\mathrm{CNR}_j(\hat{\boldsymbol{x}}) = \frac{\mathrm{crc}_j(\hat{\boldsymbol{x}})}{\sqrt{\mathsf{Var}_j(\hat{\boldsymbol{x}})}} \tag{21}$$

When assuming that the reconstructions are locally ergodic, then the CNR is equivalent to the signal-to-noise ratio (SNR). The CNR is commonly used to test the performance of PET reconstruction algorithms. In order to compare the four methods, we selected several β values and then performed image reconstruction with the mean sinogram $\bar{\boldsymbol{y}}$. In the case, q_i's were approached by: $q_i = \max\{\bar{y}_i, 1\}$ for all i. CNR values were computed using (21) where the $\mathrm{crc}_j(\hat{\boldsymbol{x}})$ and $\mathsf{Var}_j(\hat{\boldsymbol{x}})$ were calculated with the aid of theoretical expressions (10), (11) and (20). The results were plotted for the center pixel (33, 33) in Fig. 13. These curves show unique global maxima, e.g., the CNRs peak at $\beta \approx 10^{-0.9}$ for TIWT-MAP. We observe again that the TIWT-MAP slightly outperforms others because it produces the highest maximum CNR among four methods. This justifies our previous results and points out the efficiency of the proposed wavelet prior model. For QUAD-MAP and TV-MAP, the maximum CNR values are nearly equal to each other. When dealing with DWT-MAP, we found that if we use the naive choice $\hat{\boldsymbol{x}} = \boldsymbol{x}^{\mathrm{true}}$ the theoretical approximations can be applied. While this may be not correct, it could provide us some insights on how the behavior of DWT-MAP method. We traced out two β versus CNR curves for DWT-MAP at two connected pixels of interest (33, 33) and (32, 33). The results are put together in Fig. 13.

As we can see, the DWT-MAP yields two distinct CNR curves, and the maximum CNRs occur for two different β values. Therefore, the performance of DWT-MAP is highly spatially varying. As mentioned previously, this situation should attribute to the shift-varying property of DWT which makes it difficult to predict the noise behavior with the current version of theoretical approximations. In previous sections, we have already noted that besides the multiscale nature, the proposed prior model differs from the conventional MRF-based one due to an additional global coarse level constraint. Such a global penalization plays the role of noise suppression as well. To show the relevance of such *a priori* penalization, we defined a modified TIWT-MAP by setting $\alpha_M = 0$ ignoring the coefficients $\boldsymbol{\theta}_M^0$. The CNR performance of the modified TIWT-MAP was evaluated and the result is shown in Fig. 13. By comparing the maximum CNRs, we see that the performance of TIWT-MAP without the global constraint is slightly inferior to the constrained case. Nevertheless, such modified TIWT-MAP still shows better performance than QUAD-MAP and TV-MAP while the difference between two TIWT-MAP algorithms is not obvious. These results are probably due to the facts that: 1) the main contribution of a TIWT-based prior is captured by the used multiscale local features; 2) the value of α_M is relatively small (in this particular experiment $\alpha_M = 2^{-6}$), leading to less contribution of low-level approximation coefficients. However, we have to notice that a penalization on the global information perhaps results in an additional negative bias of the estimator which is undesirable for image quantification. So, how much bias can be produced by the above α_M? To get some insights to this problem, we conducted another experiment by analyzing the bias-standard deviation tradeoff of the related estimators. Fig. 14 compares the tradeoff curves based on two hot lesions of 22 mm and 17 mm diameters respectively. It can be seen that they are close to each other. This implies that the bias changes related to a global constraint are negligible, i.e., the proposed TIWT-MAP does not cause too much negative effects on image reconstruction.

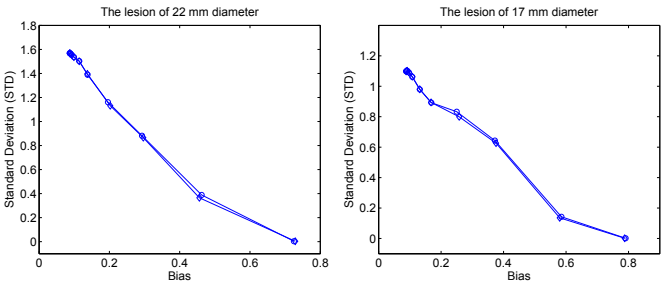
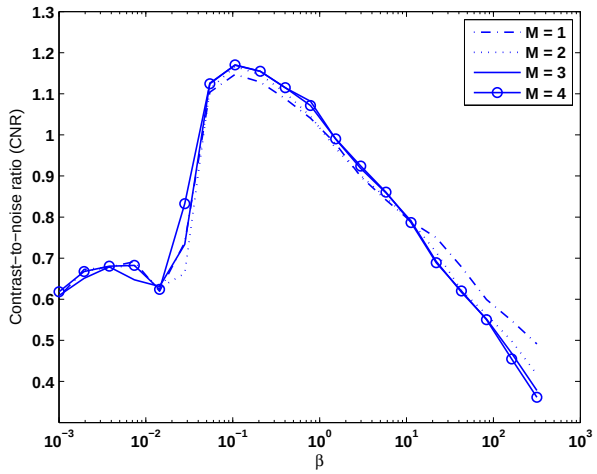


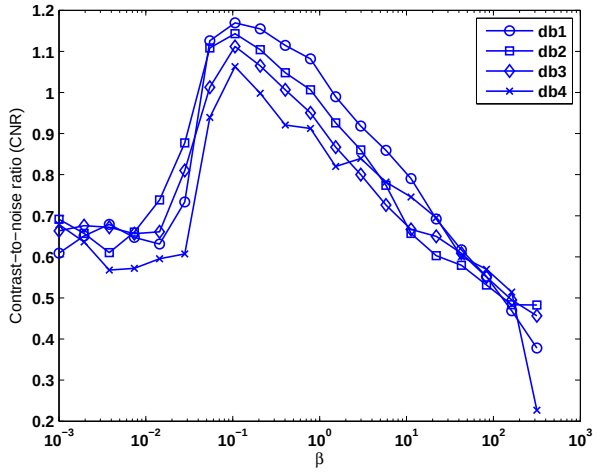
Fig. 14. The comparison of bias and standard deviation of two hot lesion estimators of (a) 22 mm and (b) 17 mm diameters using the TIWT-MAP with (circle) and without (diamond) *a priori* constraint on coefficients $\boldsymbol{\theta}_M^0$.

B. Influence of wavelet filters

For the proposed TIWT-MAP algorithm, there are at least two remaining aspects that should be considered: 1) the type of wavelet filter which characterizes the filtering behavior and 2) the used maximum scale that affects the range of frequency concentration. We first studied the latter issue, i.e., the changes of CNR performance as a function of M was examined. We used the Haar wavelet and traced out the corresponding performance curves in Fig. 15(a). We see that the maximum CNR increases when M becomes bigger. This fact exactly indicates that a multiscale prior is clearly more advantageous



(a)

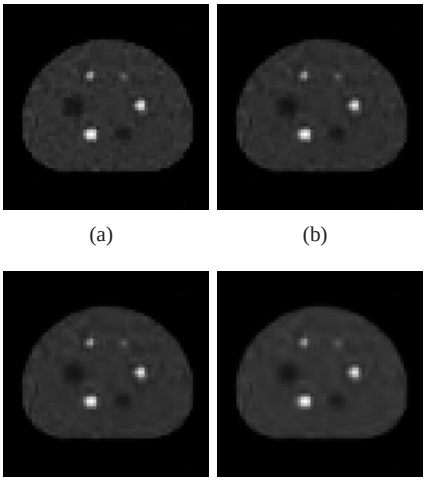


(b)

Fig. 15. (a) Relation between the maximum decomposition level M and CNR of image reconstruction using TIWT-MAP with Haar wavelets; (b) Relation between wavelet filters and CNR using TIWT-MAP with a fixed maximum decomposition level $M = 3$.

over a single scale prior when viewed through a global CNR performance criterion. On the other hand, it also reflects that a larger range of frequency concentration resulting from a multiscale prior does improve the image reconstruction of a MAP algorithm. This shows the agreement with our previous analysis on scale effects of wavelet prior. It is worth to note that the global maxima for different methods of M do not differ too much from each other, all occurring at $10^{-1.0}$. The difference between the choices $M = 3$ and $M = 4$ is not significant.

Let us investigate the effect of wavelet filters. Fig. 15(b) shows the CNR curves obtained by selecting four different wavelet filters: db1 (as well as the Haar wavelet), db2, db3, and db4 (where the number relates to the length of wavelet filters by a factor of 2). In our simulation, we have set $M = 3$, and the pixel of interest was chosen to be near the smallest 10 mm lesion. Interestingly, the Haar wavelet gives the optimal



(a) (b)

(c) (d)

Fig. 16. Image reconstructions using TIWT-MAP coupled with different wavelet filters (from (a) to (d)): db1 (as well as the Haar wavelet), db2, db3 and db4. All TIWTs were computed up to level $M = 3$, and parameters β were chosen to obtain approximately the CRC (≈ 0.3) for the pixel located at the center of phantom.

performance because it has the largest maximum CNR, while the maximum CNR decreases as increasing the wavelet filter length. This indicates that large support wavelet bases might not be suitable for small lesion detection. As an example, Fig. 16 shows image reconstructions using four types of wavelet filter. Here, the parameter β has been set in such a way that all MAPs produce nearly the same CRC values. Note that smoother edges are obtained using db4 wavelet filters. The reason is that the local impulse response has been highly penalized by the large support wavelet prior model. In fact, a similar result has been reported by Qi and Leahy [34] where the 3rd order thin plate model was shown to have poorer detectability than lower order membrane models.

IV. Discussion and Conclusion

In PET imaging, MAP methods are often introduced to improve the noise behavior of image reconstruction. The present work describes a MAP method (i.e., TIWT-MAP) using a multiscale wavelet prior. This prior, based on the heavy-tailed behavior of wavelet coefficients, was constructed from the translation invariant wavelet transform. We have discussed the design of TIWT prior and the way to choose weighting parameters based on the signal energy. The image reconstruction was performed in the spatial domain (rather than in wavelet domain) with a block sequential EM-like iteration algorithm.

Quantifying a MAP algorithm for PET image reconstruction usually can be carried out in terms of resolution and variance properties. While the TIWT-MAP estimate is nonlinear and has no closed-form expression, this still can be done locally with the aid of classical theoretical approximates, such as local impulse response function and variance. The local shift invariance provides another useful way to explore approximately the characteristics of a MAP algorithm through the local behavior of the Hessian of the used prior function. Under

some reasonable assumptions, we have found that the Hessian kernel at given pixel can be expressed into a combination of linear filters. These filters, including combination coefficients, are key components that characterize a prior model and differentiate one MAP algorithm from another.

For the proposed multiscale TIWT prior, the local behavior of the Hessian depends on what kind of wavelet filters and how many decomposition levels we use. Theoretically, we have observed that the TIWT prior would be more efficient as we increase the decomposition scale (i.e., larger than one). This is because the more scales used, the more the signal content would be captured. Empirical results have verified this prediction. However, the performance of TIWT-MAP seems bounded when we increase the level of wavelet transform. Additionally, the global *a priori* constraint on low-level approximation coefficients is of relevance and provides a slight improvement in CNR while less negative effects in mean reconstruction. A natural question follows: is there optimal α_M that tradeoffs the maximum of CNR and the bias of mean? To obtain such α_M is obviously interesting but is beyond the scope of this paper.

Another point worth to mention is that the large support wavelet filters may not be suitable for sharp edge-preserving. This is because large support wavelets usually have stronger high-pass effects (and sharper transition region) which penalize largely higher frequency components of the local impulse response. On the other hand, this also indicates that the Haar wavelets would be the favorite choice since it always has the smallest support. These theoretical considerations have been confirmed by our simulation studies.

It should be underlined that theoretical prediction for the noise properties of a MAP using nonquadratic prior is often complicated due to its highly nonlinear nature. Our theoretical prediction is mainly conducted with formulations (10) and (11). The accuracy of the local linearized impulse response and the covariance depends on the resolution and noise level in reconstructed images because of the approximations that are used. The main difficulty for prediction is perhaps that the noise level in reconstruction can be large (due to the ill-posed nature of tomography reconstruction) so that the accuracy of (10) and (11) cannot be ensured any more (in particular for pixels around sharp changes). The Hessian is especially the key to such approximation, and efficient calculation have recently been suggested by [35] and [47]. As we have already mentioned in experiments, our reconstructions were obtained from the assumed mean sinogram instead of noisy realizations since we only concern to provide some useful insights. Therefore, the noise in final reconstruction is usually small especially when large beta is used. This strategy can undoubtedly lead to better prediction than that using the noisy realizations. However, when beta is too small, there is still discrepancy between theoretical predictions and experimental simulations due to the sensitiveness of approximation.

Quantitative results show that TIWT-MAP performs better than QUAD-MAP and TV-MAP. It provides relatively higher maximum contrast-to-noise ratio and thus better lesion contrast recovery. DWT-MAP does not outperform TIWT-MAP in the same kind of application. Also our simulation results reveal that the resolution and noise properties of DWT based MAP reconstruction can be complicated. In spite of this, the DWT-MAP has its own advantages, e.g., the fast algorithm and low computational complexity. As a result, further studies on such algorithm are deserved.

The current theoretical evaluation of resolution and noise properties is limited to mean measurement data. Verification on their accuracy with one single noisy realization should be conducted. The computation would be intensive for large data sets, so we plan to design wavelet-based simplified theoretical expressions that allow fast calculation. The CNR utilizing single pixel variance and contrast recovery coefficient in task of lesion detection does not include the correlation from nearby pixels. Therefore, a further study based on criteria involving pixel correlation such as the channelized Hotelling observer (CHO) may be better to understand the performance of the proposed TIWT-MAP method. We will also investigate higher dimensional wavelet-based PET image reconstruction, and will evaluate the proposed method using real data.

ACKNOWLEDGEMENT

The authors would like to thank Prof. J. A. Fessler for providing free matlab routines to compute the NU2–2000 NEMA phantom. They also thank the reviewers for their valuable comments that helped to significantly improve the manuscript.

APPENDIX

Let $\hat{\boldsymbol{\theta}}_m^d(\omega_1,\omega_2)$, $d = 0,\ldots,3$, be the two-dimensional Fourier transform of $\boldsymbol{\theta}_m^d$ where ω_1 and ω_2 represent the frequency components. The Fourier transform of h_0 and g_0 are denoted by $\hat{h}_0(\omega)$ and $\hat{g}_0(\omega)$. Then, the one-scale TIWT decomposition can be expressed in terms of Fourier transforms

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_m^0(\omega_1,\omega_2) &= \hat{\boldsymbol{\theta}}_{m-1}^0(\omega_1,\omega_2)\hat{h}_0(2^{m-1}\omega_1)\hat{h}_0(2^{m-1}\omega_2)\\
\hat{\boldsymbol{\theta}}_m^1(\omega_1,\omega_2) &= \hat{\boldsymbol{\theta}}_{m-1}^0(\omega_1,\omega_2)\hat{h}_0(2^{m-1}\omega_1)\hat{g}_0(2^{m-1}\omega_2)\\
\hat{\boldsymbol{\theta}}_m^2(\omega_1,\omega_2) &= \hat{\boldsymbol{\theta}}_{m-1}^0(\omega_1,\omega_2)\hat{g}_0(2^{m-1}\omega_1)\hat{h}_0(2^{m-1}\omega_2)\\
\hat{\boldsymbol{\theta}}_m^3(\omega_1,\omega_2) &= \hat{\boldsymbol{\theta}}_{m-1}^0(\omega_1,\omega_2)\hat{g}_0(2^{m-1}\omega_1)\hat{g}_0(2^{m-1}\omega_2)
\end{aligned}$$

for all $m = 1,\ldots,M$. It follows immediately that

$$\begin{aligned}
&|\hat{\boldsymbol{\theta}}_m^0(\omega_1,\omega_2)|^2 + |\hat{\boldsymbol{\theta}}_m^1(\omega_1,\omega_2)|^2\\
&\quad + |\hat{\boldsymbol{\theta}}_m^2(\omega_1,\omega_2)|^2 + |\hat{\boldsymbol{\theta}}_m^3(\omega_1,\omega_2)|^2\\
&\quad = |\hat{\boldsymbol{\theta}}_{m-1}^0(\omega_1,\omega_2)|^2\Big\{(|\hat{h}_0(2^{m-1}\omega_1)|^2 + |\hat{g}_0(2^{m-1}\omega_1)|^2)\\
&\qquad \times (|\hat{h}_0(2^{m-1}\omega_2)|^2 + |\hat{g}_0(2^{m-1}\omega_2)|^2)\Big\}.
\end{aligned} \tag{22}$$

For wavelet filters, we may have

$$|\hat{h}_0(\omega)|^2 + |\hat{g}_0(\omega)|^2 = C, \quad \forall\omega \in [-\pi,\pi] \tag{23}$$

where C is a constant. This condition holds for many wavelet filters, e.g., for the commonly used conjugate mirror filters, we have $C = 2$. Substituting it into (22), we obtain

$$\begin{aligned}
|\hat{\boldsymbol{\theta}}_{m-1}^0(\omega_1,\omega_2)|^2 = \frac{1}{C^2}\Big\{&|\hat{\boldsymbol{\theta}}_m^0(\omega_1,\omega_2)|^2 + |\hat{\boldsymbol{\theta}}_m^1(\omega_1,\omega_2)|^2\\
&+ |\hat{\boldsymbol{\theta}}_m^2(\omega_1,\omega_2)|^2 + |\hat{\boldsymbol{\theta}}_m^3(\omega_1,\omega_2)|^2\Big\}.
\end{aligned}$$

By integrating both sides of the above equation and then applying the Parseval's theorem, we can approach the two-scale signal energy relationship.

REFERENCES

[1] Z. Wu, G. T. Herman and J. A. Browne, "Edge preserving reconstruction using adaptive smoothing in wavelet domain," in *Proc. IEEE Nucl. Sci. Symp. Med. Im. Conf.*, vol. 3, 1993, pp. 1917–1921.

[2] A. H. Delaney and Y. Bresler Y, "Multiresolution tomographic reconstruction using wavelets," *IEEE Trans. Image Process.*, vol. 7, pp. 799–813, 1995.

[3] M. Bhatia, W. C. Karl and A. S. Willsky, "A wavelet-based method for multiscale tomographic reconstruction," *IEEE Trans. Med. Imaging*, vol. 15, pp. 92–101, 1996.

[4] R. D. Nowak and E. D. Kolaczyk, "A statistical multiscale framework for Poisson inverse problem," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1811–1825, 2000.

[5] A. Raheja and A. P. Dhawan, "Wavelet based multiresolution expectation maximization image reconstruction algorithm for positron emission tomography," *Computerized Medical Imaging and Graphics*, vol. 24, pp. 359–376, 2000.

[6] Y. Choi Y, J. Y. Koo and N. Y. Lee, "Image reconstruction using the wavelet transform for positron emission tomography," *IEEE Trans. Med. Imaging*, vol. 20, pp. 1188–1193, 2001.

[7] N. Y. Lee and B. J. Lucier, "Wavelet methods for inverting the Radon transform with noisy data," *IEEE Trans. Image Process.*, vol. 10, pp. 79–94, 2001.

[8] C. De Mol and M. Defrise, "A note on wavelet-based inversion methods," *Inverse Problems, Image Analysis and Medical Imaging Contemporary Mathematics* vol. 313, ed M Z Nashed and O Scherzer Rhode Island: American Mathematical Society, 2002.

[9] T. Frese, C. A. Bouman and K. Sauer, "Adaptive wavelet graph model for Bayesian tomographic reconstruction," *IEEE Trans. Image Process.*, vol. 11, pp. 756–770, 2002.

[10] J. Kalifa, A. Laine and P. D. Esser, "Regularization in tomographic reconstruction using thresholding estimators," *IEEE Trans. Med. Imaging*, vol. 22, pp. 351–359, 2003.

[11] N. Y. Lee and Y. A. Choi, "A modified OSEM algorithm for PET reconstruction using wavelet processing," *Computer Methods and Programs in Biomedicine*, vol. 80, pp. 236–245, 2005.

[12] M. Rantala, S. Vänskä, S. Järvenpää, M. Kalke, M. Lassas, J. Moberg and S. Siltanen, "Wavelet-based reconstruction for limited-angle X-ray tomography," *IEEE Trans. Med. Imaging*, vol. 25, pp. 210–217, 2006.

[13] J. Zhou, J. L. Coatrieux, A. Bousse, H. Z. Shu and L. M. Luo, "A Bayesian MAP-EM algorithm for PET image reconstruction using wavelet transform," *IEEE Nucl. Sci.*, vol. 51, pp. 1660–1669, 2007.

[14] D. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *Applied and Computational Harmonic Analysis*, vol. 2, pp. 101–126, 1995.

[15] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-gaussian and complexity priors," *IEEE Trans. Inform. Theory.*, vol. 38, pp. 719–747, 1998.

[16] M. Belge, M. Kilmer and E. Miller, "Wavelet domain image restoration with adaptive edge-preserving regularization," *IEEE Trans. Image Process.*, vol. 9, pp. 597–608, 2000.

[17] A. Antoniadis, D. Leporini and J. C. Pesquet, "Wavelet thresholding for some classes of non-Gaussian noise," *Statist. Neerlandica*, vol. 56, pp. 434–453, 2002.

[18] I. Daubechies, M. Defrise and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004

[19] I. Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia, 1992.

[20] J. C. Pesquet, H. Krim and H. Carfantan, "Time-invariant orthonormal wavelet presentation," *IEEE Trans. Signal Processing*, vol. 44, pp. 1964–1970, 1996.

[21] S. G. Mallat, *A Wavelet Tour of Signal Processing, second edition*, Academic Press, San Diego, 1999.

[22] L. Shen, M. Papadakis, I. A. Kakadiaris, I. Konstantinidis, D. Kouri, D. Hoffman, "Image denoising using a tight frame," *IEEE Trans. Image Process.*, vol. 15, pp. 1254–1263, 2006.

[23] C. Chaux, P. L. Combettes, J. C. Pesquet and V. R. Wajs, "A variation formulation for frame-based inverse problems," *Inverse Problems*, vol. 23, pp. 1495–1518, 2007.

[24] A. R. De Pierro and M. E. Yamagishi, "Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography," *IEEE Trans. Med. Imaging*, vol. 20, pp. 280–288, 2001.

[25] M. Holschneider, R. Kronland-Martinet, J. Morlet and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," *Wavelets, Time-Frequency Methods and Phase Space*, Springer-Verlag, Berlin, 1989.

[26] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," *Wavelet and Statistics, Lecture Notes in Statistics*, Berlin Springer, 1995.

[27] M. Lang, H. Guo, J. Odegard, C. S. Burrus and R. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Process. Lett.*, vol. 3, pp. 10–12, 1995.

[28] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," *Wavelet and Statistics, Lecture Notes in Statistics*, Springer, New York, 1995.

[29] S. G. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform. Theory.*, vol. 38, pp. 617–643, 1992.

[30] H. Choi and R. G. Baraniuk, "Wavelet statistical models and Besov spaces," *Proc. SPIE*, vol. 3813, pp. 489–501, 1999.

[31] H. Choi H and R. G. Baraniuk, "Multiple wavelet basis image denoising using besov ball projections," *IEEE Signal Process. Lett.*, vol. 9, pp. 717–720, 2004.

[32] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography," *IEEE Trans. Image Process.*, vol. 5, pp. 493–506, 1996.

[33] J. A. Fessler and W. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs," *IEEE Image Process.* vol. 5, pp. 1346–1358, 1996.

[34] J. Qi and R. M. Leahy, "A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data," *IEEE Trans. Med. Imaging*, vol. 18, pp. 293–305, 1999.

[35] J. Qi and R. M. Leahy, "Resolution and noise properties of MAP reconstruction for fully 3D PET," *IEEE Trans. Med. Imaging*, vol. 19, pp. 493–506, 2000.

[36] J. A. Fessler, "Penalized weighted least squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imaging*, vol. 13, pp. 290–300, 1994.

[37] Q. Li, E. Asma, J. Qi, J. R. Bading and R. M. Leahy, "Accurate estimate of the fisher information matrix for the PET image reconstruction problem," *IEEE Trans. Med. Imaging*, vol. 23, pp. 1057–1064, 2004.

[38] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Trans. Med. Imaging* vol. 19, pp. 601–615, 2000.

[39] J. W. Stayman and J. A. Fessler, "Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems," *IEEE Trans. Med. Imaging*, vol. 23, pp. 269–284, 2004.

[40] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Image Process.*, vol. 8, pp. 688–699, 1999.

[41] J. A. Fessler, "Analytical approach to regularization design for isotropic spatial resolution," in *Proc. IEEE Nucl. Sci. Symp. Med. Im. Conf.*, vol. 3, 2003, pp. 2022–2026.

[42] J. Browne and A. R. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography," *IEEE Trans. Med. Imaging*, vol. 15, pp. 687–699, 1996.

[43] S. J. Kim, K. Kih, M. Lustig, S. Boyd and D. Gorinevsky, "An interior point method for large scale ell1 regularized least squares," *IEEE J. Select. Top. in Sig. Proc.*, vol. 1, no. 4, pp. 606–617, 2007.

[44] National Electrical Manufacturers Association (NEMA) 2000 Performance measurements of positron emission tomographs *NEMA Standard Publication* NU2–2000 (Washington, DC: NEMA)

[45] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Trans. Image Process.*, vol. 7, pp. 813–824, 1998.

[46] J. Qi, R. M. Leahy, C. Hsu, T. H. Farquhar and S. R. Chery, "Fully 3D Bayesian image reconstruction for the ECAT EXACT HR+," *IEEE Trans. Nucl. Sci.*, vol. 45, pp. 1096–1103, 1998.

[47] S. Ahn and R. M. Leahy, "Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for PET," *IEEE Trans. Medical Imaging*, vol. 27, pp. 413–424, 2008.